Compressive Principal Component Pursuit

Yi Ma

Visual Computing Group
Microsoft Research Asia, Beijing

ECE Department
University of Illinois, Urbana

John Wright
(EE, Columbia Univ.)

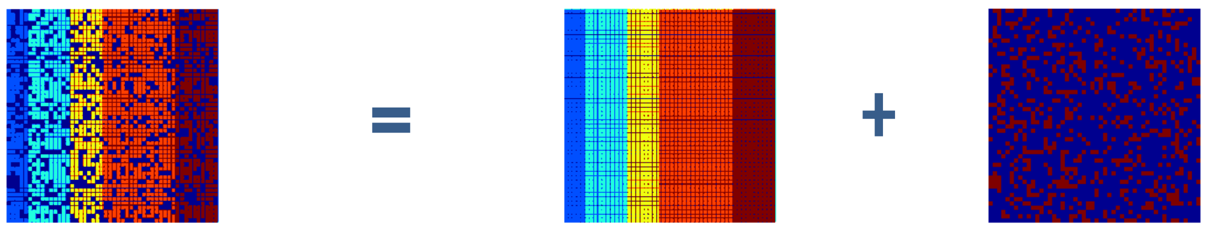
Arvind Ganesh
(IBM Research, India)

Kerui Min
(ECE, UIUC)

Structured Signal Recovery

Robust PCA (?) and beyond…

Assumption: matrix-valued signal $\boldsymbol{D} \in \mathbb{R}^{m \times n}$ can be modeled as $\boldsymbol{D} = \boldsymbol{L} + \boldsymbol{S}$ with $\boldsymbol{L}$ *low-rank,* $\boldsymbol{S}$ *sparse*.



Good model for many natural signals:

Structured Signal Recovery

Robust PCA (?) and beyond…

Convex optimization recovers $(\boldsymbol{L}_0, \boldsymbol{S}_0)$, under conditions*:

$$(\boldsymbol{L}_0, \boldsymbol{S}_0) = \arg \min \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{D}.$$

* $\boldsymbol{L}_0$ rank-r, μ-incoherent, $\operatorname{supp}(\boldsymbol{S}_0) \sim \operatorname{Ber}(\rho)$, $\rho < \rho_0$,

and $r < \frac{cn}{\mu \log^2 m}$.

Above result: [Candès, Li, Ma, Wright, JACM'11]

Recovery program: [Chandrasekaran, Sanghavi, Parillo, Willsky, SIOPT'11]

Structured Signal Recovery

Robust PCA (?) and beyond…

Convex optimization recovers $(\boldsymbol{L}_0, \boldsymbol{S}_0)$, under conditions*:

$$(\boldsymbol{L}_0, \boldsymbol{S}_0) = \arg\ \min\ \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1$$
$$\text{s.t. } \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{D}.$$

$\boldsymbol{L}_0$ rank-r, μ-**incoherent**: $\boldsymbol{L}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$,

with
$$\|\boldsymbol{U}\|_{2,\infty} \le \sqrt{\mu r/m}$$
$$\|\boldsymbol{V}\|_{2,\infty} \le \sqrt{\mu r/n}$$
$$\|\boldsymbol{U}\boldsymbol{V}^*\|_\infty \le \sqrt{\mu r/mn}$$

), $\rho < \rho_0$,

A

Recovery program: [Chandrasekaran, Sanghavi, Parillo, Willsky, SIOPT'11]

Structured Signal Recovery

Robust PCA (?) and beyond…

This talk:

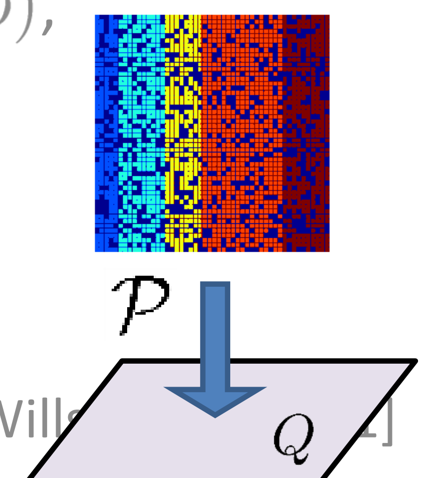
What if we only observe a few measurements

$\mathcal{P}_Q[\boldsymbol{D}]$? ($Q \subseteq \mathbb{R}^{m\times n}$ linear, low-dim)

Can we still recover $(\boldsymbol{L}, \boldsymbol{S})$?

*$\boldsymbol{L}_0$ rank-r, μ-incoherent, $\mathrm{supp}(\boldsymbol{S}_0) \sim \mathrm{Ber}(\rho)$,

and $r < \frac{cn}{\mu \log^2 m}$.



Above result: [Candès, Li, Ma, Wright, JACM’11]

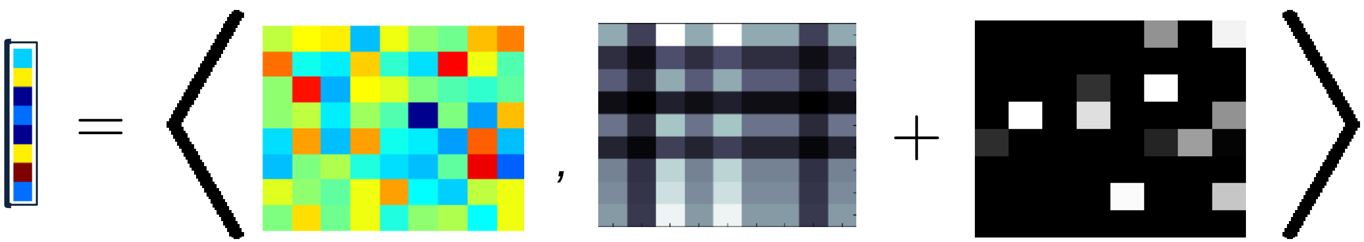
Recovery program: [Chandrasekaran, Sanghavi, Parillo, Wil 1]

Motivating Applications: Compressed Sensing

Compressive Imaging of Videos and Hyperspectral Images:

SpaRCS [Waters, Sankarayanan, and Baraniuk, NIPS'11]

Problem: Given $\boldsymbol{Y} = \mathcal{P}_Q(\boldsymbol{L}_0 + \boldsymbol{S}_0)$, recover $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$.



$\boldsymbol{Y} \in \mathbb{R}^q$ $\quad \boldsymbol{Q}_i \quad$ $\boldsymbol{L}_0 + \boldsymbol{S}_0 \quad$ $i = 1, \ldots, q$

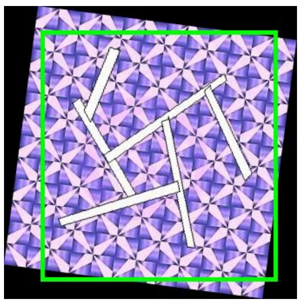
Motivating Applications: Domain Transformations

Applications to Transformation Invariant Low-rank Textures:
TILT [Zhang, Xiao, Ganesh, and Ma, IJCV 2011]

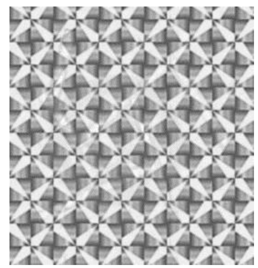
$\boldsymbol{D}$ – corrupted & distorted observation

$\boldsymbol{L}$ – aligned low-rank signals

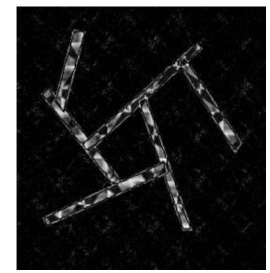
$\boldsymbol{S}$ – sparse errors



$\circ\, \tau \; =$



$+$



Problem: Given $\boldsymbol{D} \circ \tau \;=\; \boldsymbol{L}_0 + \boldsymbol{S}_0$, recover τ, $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$.

Parametric domain deformations
(rigid, affine, projective, radial distortion, 3D shapes…)

Motivating Applications: Structured Texture Inpainting

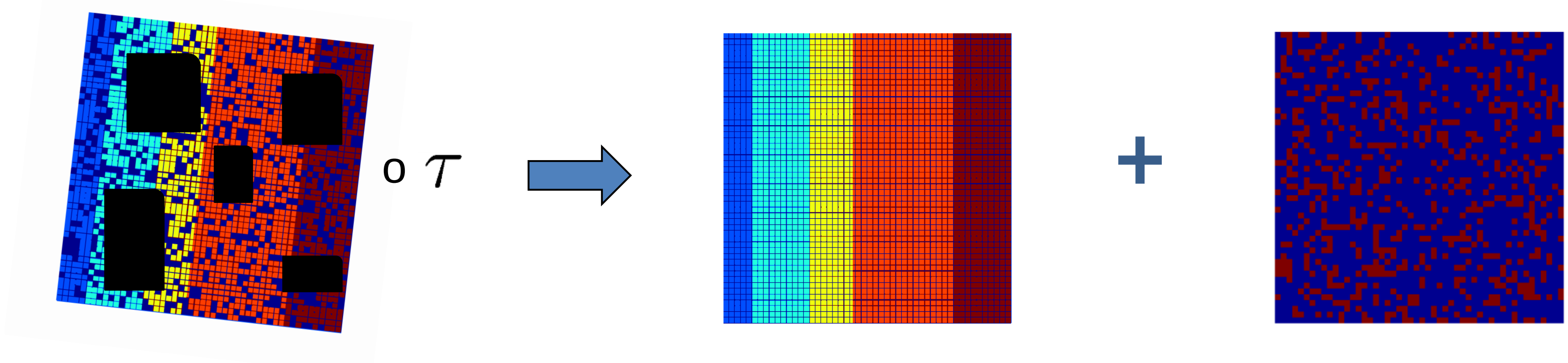
Automatic Low-rank and Sparse **Texture Repairing**:
[Xiao, Ren, Zhang, and Ma, ECCV 2012]

Problem: Given $\boldsymbol{Y} \circ \tau = \mathcal{P}_\Omega(\boldsymbol{L}_0 + \boldsymbol{S}_0)$, recover $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$.

Corrupted and Incomplete Samples $\boldsymbol{Y}$

Low-rank Texture $\boldsymbol{L}$

Sparse Errors $\boldsymbol{S}$



Motivating Applications: Domain Transformations

Robust Alignment of Multiple Images via Sparse and Low-rank Decomposition:
RASL [Peng, Ganesh, Wright, and Ma, CVPR 2010]

$\boldsymbol{D}$ – corrupted & misaligned observation

$\boldsymbol{L}$ – aligned low-rank signals

$\boldsymbol{S}$ – sparse errors

$\circ\, \tau \;=$

$+$

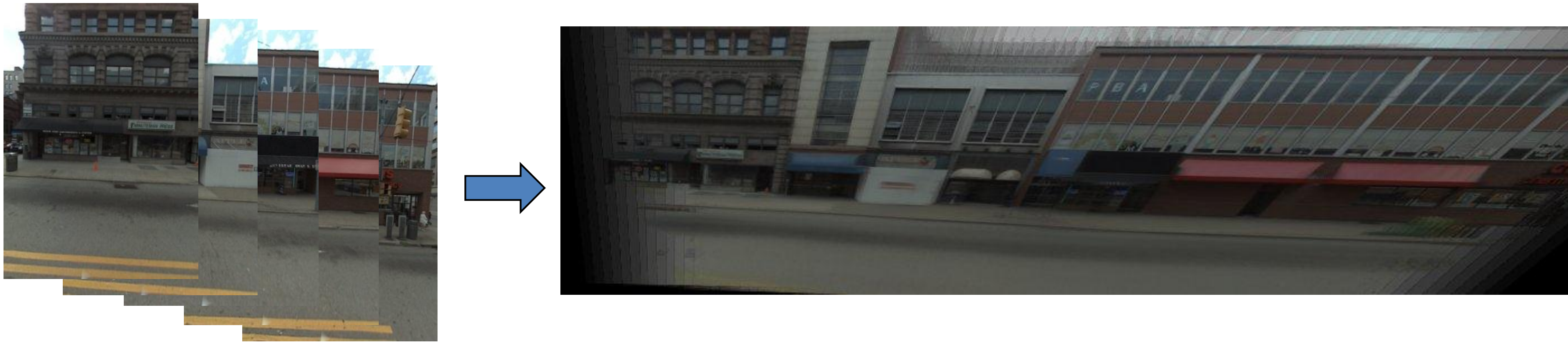
Problem: Given $\boldsymbol{D} \circ \tau \;=\; \boldsymbol{L}_0 + \boldsymbol{S}_0,$ recover τ, $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$.

Linearized subproblem: $\boldsymbol{D} \circ \tau_0 + \mathcal{J} \Delta \tau = \boldsymbol{L} + \boldsymbol{S}$

$$\mathcal{P}_Q[\boldsymbol{D} \circ \tau_0] = \mathcal{P}_Q[\boldsymbol{L} + \boldsymbol{S}] \qquad Q = \mathrm{range}(\mathcal{J})^{\perp}$$

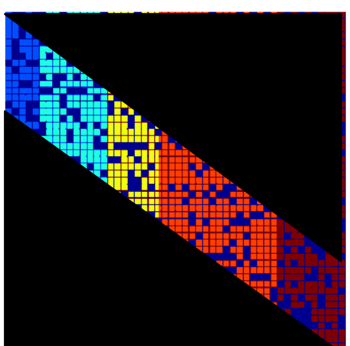
Motivating Applications: Video Panorama

Automatic Low-rank Panorama from Street View Videos:
[Zhou, Min, and Ma, submitted to NIPS 2012]



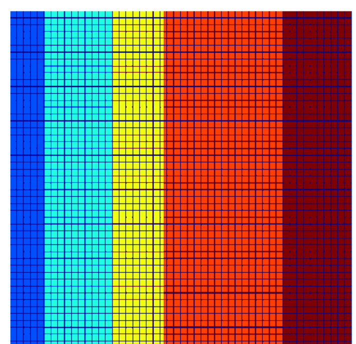
Problem: Given $\boldsymbol{Y} \circ \tau = \mathcal{P}_\Omega(\boldsymbol{L}_0 + \boldsymbol{S}_0)$, recover $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$.

Corrupted and Incomplete Samples $\boldsymbol{Y}$



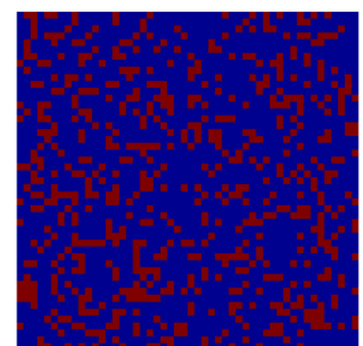
$\circ \; \tau$

Low-rank Texture $\boldsymbol{L}$



+

Sparse Errors $\boldsymbol{S}$



Problem Formulation

Compressive Principal Component Pursuit

We observe: $\mathcal{P}_Q[\boldsymbol{D}] = \mathcal{P}_Q[\boldsymbol{L} + \boldsymbol{S}]$

Natural heuristic:

$$\min \ \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \mathcal{P}_Q[\boldsymbol{L} + \boldsymbol{S}] = \mathcal{P}_Q[\boldsymbol{D}]$$

Analytical assumption:

Q is a *random subspace* (Haar measure).

Main Result

Correct Recovery from Near-Minimal Observations

Theorem 1 (Wright, Ganesh, Min, Ma) *Let* $\boldsymbol{L}, \boldsymbol{S} \in \mathbb{R}^{m \times n}$, $m \geq n$, $\boldsymbol{L}$ *μ-incoherent,* $\mathrm{sign}(\boldsymbol{S})$ *iid* $\mathrm{Bernoulli}(\rho) - \mathrm{Rademacher}$, $\rho < \rho_0$,

$$r \;=\; \mathrm{rank}(\boldsymbol{L}) \;<\; \frac{cn}{\mu \log^2 m}.$$

Then if

$$\dim(Q) \;\geq\; C\,(\rho mn + mr) \log^2 m,$$

with high probability convex programming exactly recovers $(\boldsymbol{L}, \boldsymbol{S})$.

Interpretation:

Success when $\#\mathrm{obs.} \;\geq\; \#\mathrm{dof}(\boldsymbol{L}, \boldsymbol{S}) \times O(\log^2 m)$*!*

Proof Strategy

Find a **Lagrange multiplier** $\boldsymbol{\Lambda}$ that certifies optimality of $(\boldsymbol{L}, \boldsymbol{S})$

KKT Conditions, **Compressive Principal Component Pursuit:**

$$\min \; \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \mathcal{P}_Q[\boldsymbol{L} + \boldsymbol{S}] = \mathcal{P}_Q[\boldsymbol{D}]$$

$(\boldsymbol{L}, \boldsymbol{S})$ optimal iff $\exists\, \boldsymbol{\Lambda}$ such that $\begin{cases} \boldsymbol{\Lambda} \in \partial \|\cdot\|_*(\boldsymbol{L}) \\ \boldsymbol{\Lambda} \in \lambda \partial \|\cdot\|_1(\boldsymbol{S}) \\ \boldsymbol{\Lambda} \in Q \end{cases}$

Proof Strategy

Find a **Lagrange multiplier** $\boldsymbol{\Lambda}$ that certifies optimality of $(\boldsymbol{L}, \boldsymbol{S})$

KKT Conditions, ~~Compressive~~ **Principal Component Pursuit:**

$$\min \ \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{D}$$

$(\boldsymbol{L}, \boldsymbol{S})$ optimal iff $\exists\, \boldsymbol{\Lambda}$ such that $\begin{cases} \boldsymbol{\Lambda} \in \partial \|\cdot\|_*(\boldsymbol{L}) \\ \boldsymbol{\Lambda} \in \lambda \partial \|\cdot\|_1(\boldsymbol{S}) \\ \cancel{\boldsymbol{\Lambda} \in Q} \end{cases}$

Proof Strategy

Find a **Lagrange multiplier** $\boldsymbol{\Lambda}$ that certifies optimality of $(\boldsymbol{L}, \boldsymbol{S})$

KKT Conditions, ~~Compressive~~ **Principal Component Pursuit:**

$$\min \; \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \quad \text{s.t.} \quad \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{D}$$

$(\boldsymbol{L}, \boldsymbol{S})$ optimal iff $\exists\, \boldsymbol{\Lambda}$ such that $\begin{cases} \boldsymbol{\Lambda} \in \partial \| \cdot \|_* (\boldsymbol{L}) \\ \boldsymbol{\Lambda} \in \lambda \partial \| \cdot \|_1 (\boldsymbol{S}) \\ \cancel{\boldsymbol{\Lambda} \in Q} \end{cases}$

Good approximate constructions in the literature.
[CLMW'11, HKZ, etc.].

Proof Strategy

Upgrade a certificate to a *compressive* certificate. More generally:

Multiple structure decomposition:

$$\min \sum_i \lambda_i \|\boldsymbol{X}_i\|_{\diamond_i} \quad \text{s.t.} \quad \sum_i \boldsymbol{X}_i = \boldsymbol{D}$$

Compressive multiple structure decomposition:

$$\min \sum_i \lambda_i \|\boldsymbol{X}_i\|_{\diamond_i} \quad \text{s.t.} \quad \mathcal{P}_Q[\sum_i \boldsymbol{X}_i] = \mathcal{P}_Q[\boldsymbol{D}]$$

Examples: **PCP** [CLMW'11], **outlier pursuit** [Xu+Caramanis+Sanghavi], **morphological component analysis** [Bobin et. al.], many more …

Decomposable Norms

Say a norm $\|\cdot\|$ is **decomposable** at $\boldsymbol{X}$ if there is a subspace T and $\boldsymbol{S} \in T$ s.t.:

$$\partial\|\cdot\|(\boldsymbol{X}) = \{\boldsymbol{\Lambda} \mid \mathcal{P}_T \boldsymbol{\Lambda} = \boldsymbol{S},\ \|\mathcal{P}_{T^\perp} \boldsymbol{\Lambda}\|^* \leq 1\}.$$

Examples:

$$\|\cdot\|_1 \ :\ T = \mathrm{supp}(\boldsymbol{X}),\quad \boldsymbol{S} = \mathrm{sign}(\boldsymbol{X})$$

$$\|\cdot\|_* \ :\ \boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,\ T = \{\boldsymbol{U}\boldsymbol{P} + \boldsymbol{Q}\boldsymbol{V}^*\},\quad \boldsymbol{S} = \boldsymbol{U}\boldsymbol{V}^*$$

Sum-of-column norms, many structured regularizers [Bach’11].

Decomposability: [Candès+Recht’11], see also [Negahban et. al. ’10].

Relax!

Consider the multiple-term decomposition problem

$$\min \ \sum_i \lambda_i \|\boldsymbol{X}_i\|_{\diamond_i} \quad \text{s.t.} \quad \sum_i \boldsymbol{X}_i = \boldsymbol{D}$$

KKT demands $\mathcal{P}_{T_i}\boldsymbol{\Lambda} = \boldsymbol{S}, \ \|\mathcal{P}_{T_i^\perp}\boldsymbol{\Lambda}\|_{\diamond_i} \le \lambda_i.$

Call $\boldsymbol{\Lambda}$ an (α, β) **inexact certificate** if

$$\|\mathcal{P}_{T_i}\boldsymbol{\Lambda} - \boldsymbol{S}\|_F \le \alpha, \quad \|\mathcal{P}_{T_i^\perp}\boldsymbol{\Lambda}\|_{\diamond_i}^* \le \beta\lambda_i.$$

For the **compressive variant**, additional constraint: $\boldsymbol{\Lambda} \in Q.$

Golfing Construction

Let $\hat{\mathbf{\Lambda}}$ be a certificate for the fully observed problem $\sum_i \boldsymbol{X}_i = \boldsymbol{D}$.

For compressive observations $\mathcal{P}_Q[\sum_i \boldsymbol{X}_i] = \mathcal{P}_Q[\boldsymbol{D}]$**,**

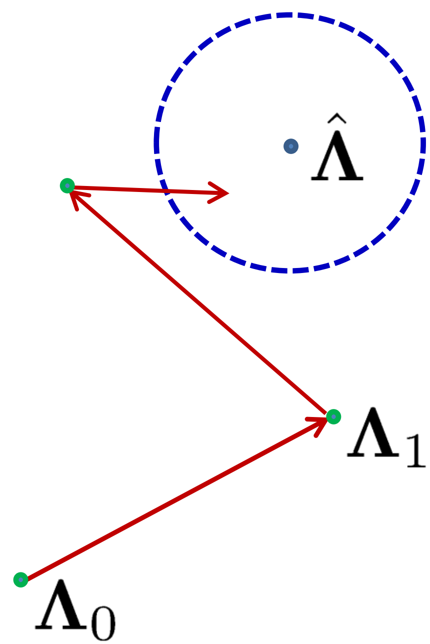
want $\mathbf{\Lambda} \approx \hat{\mathbf{\Lambda}},\ \mathbf{\Lambda} \in Q$**.**

Write $Q = \underbrace{\mathcal{R}[\mathcal{A}_1] + \cdots + \mathcal{R}[\mathcal{A}_\gamma]}$.

Independent self-adjoint operators $\mathbb{E}\mathcal{A}_i = \frac{\gamma}{mn}\mathcal{I}$

Set $\mathbf{\Lambda}_i = \mathbf{\Lambda}_{i-1} + \frac{mn}{\gamma}\mathcal{A}_i[\underbrace{\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}_{i-1}}]$.

Error at iteration i



General Upgrade Theorem

Theorem 2 *Let* $\hat{\mathbf{\Lambda}}$ *an* (α, β) *inexact certificate for the fully observed problem. Then if*

$$\dim(Q) \;\geq\; C \cdot \dim(T_1 + \cdots + T_k) \cdot \log m,$$

whp, there is an (α', β') *certificate* $\mathbf{\Lambda}$ *for the compressive problem, with*

$$\begin{aligned} \alpha' &\leq \alpha + m^{-3} \|\hat{\mathbf{\Lambda}}\|_F \\ \beta' &\leq \beta + C \max_i \frac{\nu_i + \sqrt{\log m}}{\lambda_i} \left(\frac{\|\hat{\mathbf{\Lambda}}\|_F \log m}{\dim(Q)} \right)^{1/2} \end{aligned}$$

Apply together with [CLMW'11] to obtain Theorem 1.

Main Result

Correct Recovery from Near-Minimal Observations

Theorem 1 (Wright, Ganesh, Min, Ma) *Let* $\boldsymbol{L}, \boldsymbol{S} \in \mathbb{R}^{m \times n}$, $m \geq n$, $\boldsymbol{L}$ μ*-incoherent,* $\mathrm{sign}(\boldsymbol{S})$ *iid* $\mathrm{Bernoulli}(\rho) - \mathrm{Rademacher}$, $\rho < \rho_0$,

$$r \;=\; \mathrm{rank}(\boldsymbol{L}) \;<\; \frac{cn}{\mu \log^2 m}.$$

Then if

$$\dim(Q) \;\geq\; C\,(\rho mn + mr) \log^2 m,$$

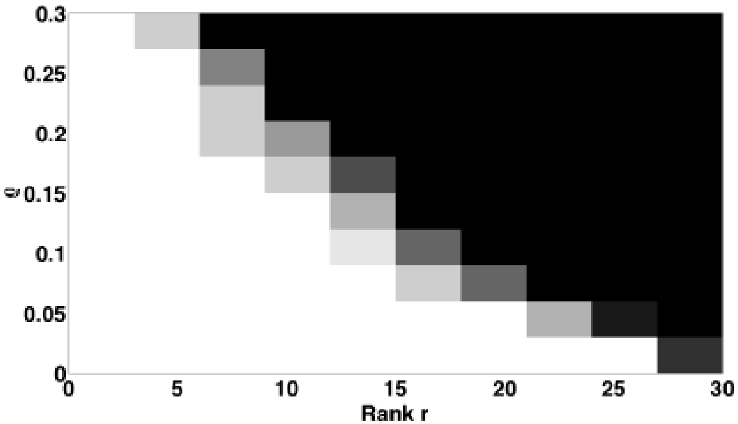
with high probability convex programming exactly recovers $(\boldsymbol{L}, \boldsymbol{S})$.

Interpretation:

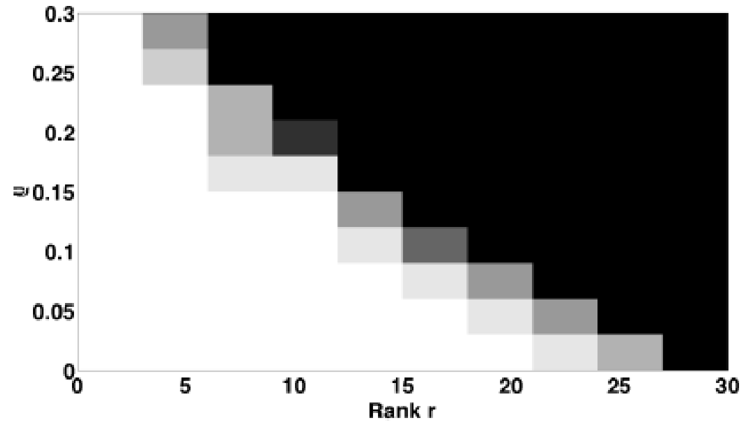
Success when $\#\mathrm{obs.} \geq \#\mathrm{dof}(\boldsymbol{L}, \boldsymbol{S}) \times O(\log^2 m)$ *!*

Numerical Examples

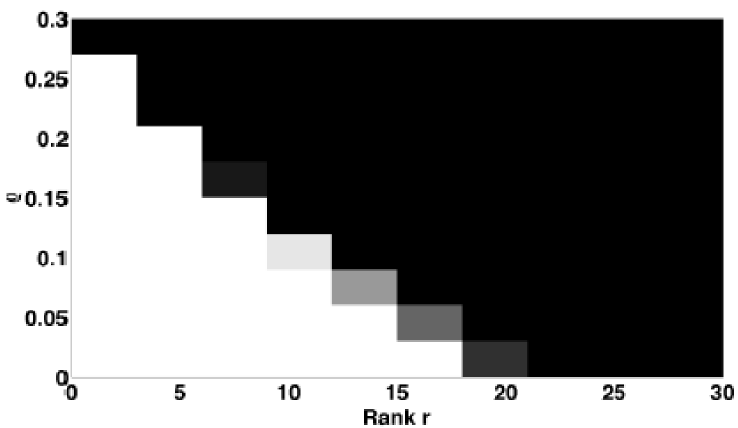
For random matrices of size 100x100:



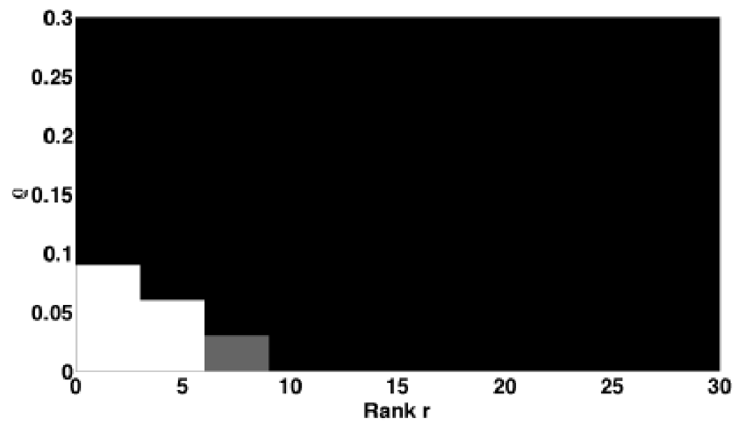
(a) $p = 0$



(b) $p = 500$



(c) $p = 2,000$



(d) $p = 5,000$

Extension I

Theorem 3 (Random Reduction). *Let $Q^\perp$ be a p-dimensional random subspac of $\mathbb{R}^{m\times n}$, $\boldsymbol{L}_0 \in \mathbb{R}^{m\times n}$, $m \geq n$ have rank r, and $\boldsymbol{S}_0$ have a Bernoulli support with error probability ρ. Then if*

$$r \;\leq\; \rho_r \frac{m}{\mu^2 \log^2(n)}, \quad p < C_p n, \quad \rho < \rho^\star$$

then with very high probability

$$(\boldsymbol{L}_0, \boldsymbol{S}_0) \;=\; \arg\min \|\boldsymbol{L}\|_* + \frac{1}{\sqrt{m}}\|\boldsymbol{S}\|_1 \quad \text{subj} \quad \mathcal{P}_Q[\boldsymbol{L}+\boldsymbol{S}] = \mathcal{P}_Q \boldsymbol{L}_0 + \boldsymbol{S}_0],$$

for some numerical constant ρ_r, C_p and $\rho^\star$, and the minimizer is unique.

PCP with Reduced Linear Measurements, Ganesh, Min, Wright, Ma, ISIT 2012.

Extension II

A subspace $S \subseteq \mathbb{R}^{m \times n}$ is ν-**coherent** if there exists an orthonormal basis $\{G_i\}$ for S satisfying $\max_i \|G_i\|^2 \leq \nu / \min\{m, n\}$.

Theorem 4 (Deterministic Reduction). *Let $Q^\perp$ be a fixed p-dimensional subspac of $\mathbb{R}^{m \times n}$, which is ν-coherent. Suppose $\boldsymbol{L}_0 \in \mathbb{R}^{m \times n}$, $m \geq n$ have rank r, and $\boldsymbol{S}_0$ have a Bernoulli support with error probability ρ. Then if*

$$r \;\leq\; \rho_r \left(\frac{m}{\alpha \nu \mu p}\right)^{1/3}, \quad \rho < \rho^\star$$

then with very high probability

$$(\boldsymbol{L}_0, \boldsymbol{S}_0) \;=\; \arg\min \|\boldsymbol{L}\|_* + \frac{1}{\sqrt{m}} \|\boldsymbol{S}\|_1 \quad \text{subj} \quad \mathcal{P}_Q[\boldsymbol{L} + \boldsymbol{S}] = \mathcal{P}_Q[\boldsymbol{L}_0 + \boldsymbol{S}_0],$$

for some numerical constant ρ_r and $\rho^\star$, and the minimizer is unique.

PCP with Reduced Linear Measurements, Ganesh, Min, Wright, Ma, ISIT 2012.

A Brief Discussion

Main results:

- Provable recovery of superpositions of structured terms (e.g., low-rank + sparse) from random measurements.

- "Upgrade" theorem for compressive variants of general convex decomposition problems.

Open issues:

- Other distributions or deterministic conditions on Q?
  Fourier-type measurements [Liu '11]? From image Jacobians?

- Noisy recovery (additive or noise folding)?

- General theory for convex decomposition problems?
  [McCoy + Tropp '12] for steps in this direction.

Compressive Principal Component Pursuit

Main References:

Compressive Principal Component Pursuit, Wright, Ganesh, Min, Ma, ISIT 2012, submitted to the IMA journal of Information and Inference.

PCP with Reduced Linear Measurements, Ganesh, Min, Wright, Ma, ISIT 2012.

Thanks to: ONR N00014-09-1-0230, NSF CCF 09-64215, NSF IIS 11-16012, and DARPA KECoM 10036-100471,

and thanks to

John Wright
(EE, Columbia Univ.)

Arvind Ganesh
(IBM Research, India)

Kerui Min
(ECE, UIUC)